

Open Research Online

The Open University's repository of research publications and other research outputs

Individual Differences and Task Behaviour

Conference or Workshop Item

How to cite:

Hall, Mark Michael and Koolen, Marijn (2016). Individual Differences and Task Behaviour. In: Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum (Balog, Krisztian; Cappellato, Linda; Ferro, Nicola and Macdonald, Craig eds.) pp. 1106–1114.

For guidance on citations see [FAQs](#).

© 2016 The Authors



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://ceur-ws.org/Vol-1609/16091106.pdf>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Individual Differences and Task Behaviour

Mark M Hall¹ and Marijn Koolen²

¹ Edge Hill University, St Helens Road, Ormskirk, L39 4QP, United Kingdom,
`Mark.Hall@edgehill.ac.uk`

² University of Amsterdam, Netherlands
`marijn.koolen@uva.nl`

Abstract. The Interactive Social Book Search track at CLEF has run the same experiment, task, and interface for two years. This provides an opportunity to study the individual differences between two separately recruited participant cohorts, rather than between sub-sets of a single cohort. Overall the results show no significant differences in how the participants used the three main stages of the interface for the two tasks. However, at the detail level there are some quite significant changes in exactly how participants use the available functionality.

Keywords: user study, interactive information retrieval, individual differences, log analysis

1 Introduction

The aim of the Interactive Social Book Search (iSBS) track at CLEF is to investigate user behaviour when faced with a collection of approximately 1.5 million books that combines both professional and user-generated content. It is now in its third year and while significant changes were made from year one to year two, the experiment, its data, interface, and tasks were kept stable between years two and three. The only major change between the two years was the participants. It is thus possible to investigate the impact of two participant cohorts without having to sub-set a single experiment based on some criteria. From this we can investigate the stability of any conclusions, both at the macro and micro levels.

The iSBS experiment [2, 7] consists of two tasks, a *non-goal* and a *goal-oriented* task. In the *non-goal* task participants were instructed to simply explore the collection until they are bored, adding any books they feel are interesting to their bookbag. In the *goal-oriented* task participants were instructed to find five books that they would like to have if they were alone on a desert island for one month.

In years one and two participants were allocated to either a baseline, faceted search interface or a novel multi-stage interface. In year three only the multi-stage interface was used. The multi-stage interface is designed to mimic the gradual narrowing of the information journey [6, 8] and is split into three stages. The initial stage (*Explore*, fig. 1a) provides a pure browsing interface for exploring the collection. The left-hand side shows a tree structure automatically

generated from the books' Amazon browse-node tags using an adapted version of the algorithm in [3]. When the user selects a tag, its child tags are shown in the tree, to allow digging down in to the tree, and on the right the books that are tagged with the tag are listed in a dense list. The user can view each book's details by clicking on the book's title.

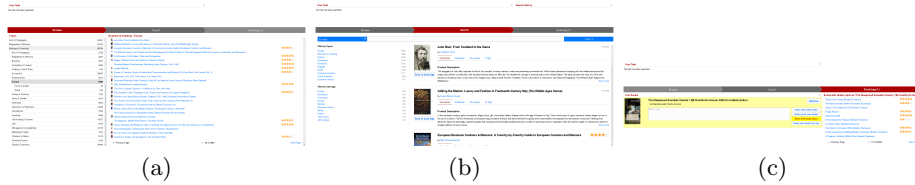


Fig. 1. Three stages of the multi-stage interface. (a) *Explore*, (b) *Search*, (c) *Book-bag*

The second stage (*Search*, fig. 1b) provides a standard faceted search interface [4], with more detail shown for each book. The third stage (*Book-bag*, fig. 1c) lets the user interact with those books that they have added to their book-bag in the *Explore* and *Search* stages. Additionally for each book the participant has access to similar books, with the similarity based on one of: authors, title, topics, or user-generated tags.

The interface, data-set, and tasks were all kept exactly the same between year two (2015) and year three (2016). There was a minor change in the experiment structure. In 2015 participants undertook both the *non-goal* and *goal-oriented* tasks (ordering balanced), while in 2016 participants were randomly assigned to one of the two tasks. This change was made to reduce the time requirements for participants. Previous analysis [1] indicates that in 2015 task order had no significant impact on use patterns, ensuring comparability with the 2016 data. Additionally, in 2016 participants could opt-in to do an additional focused task. As no comparable data is available for 2015, the additional task data is not taken into account in this analysis. The remainder of the paper will now compare the 2015 and 2016 results to investigate the impact of participant cohort differences on the observed results.

2 Explore, Query, and the Book-bag

Before looking at detailed differences in the participants' interactions with the system, it is necessary to first determine whether there are any significant high-level differences between how the participants used the system. In particular whether they used the *Explore*, *Search*, and *Book-bag* stages differently.

The experiments in both years captured very rich log data that allows for the full replication of each participant's interactions with the interface. Based on the time-stamps at which participants switched between the three stages, the amount of time they spent in each stage was calculated. As there is a large

amount of variation in how long participants spent on the system, the times were then normalised by the total session length.

Figure 2 shows the normalised time distributions for both 2015 and 2016 in the *non-goal* task. Mann-Whitney U tests showed no significant differences in the amount of time spent in each of the stages.

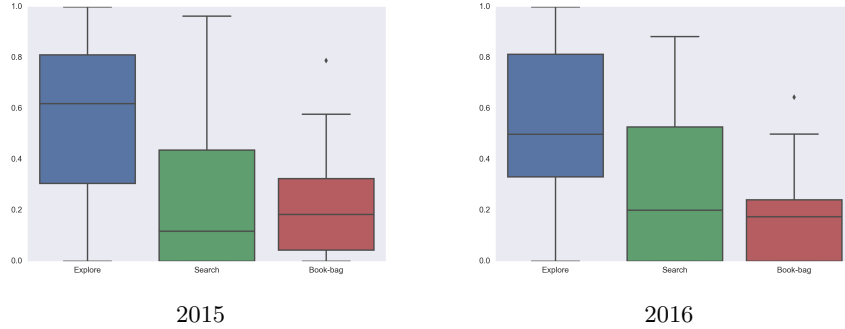


Fig. 2. Normalised time spent in each of the three stages of the multi-stage interface in the *non-goal* task.

Figure 3 shows the normalised time distributions for both 2015 and 2016 in the *goal-oriented* task. There are no significant differences between 2015 and 2016 for either the *Explore* or *Search* stages of the interface. However, in 2016 participants spent significantly more time in the *Book-bag*, compared to 2015 (Mann-Whitney’s U $p < 0.01$). In 2015 participants spent on average 15% of their time in the *Book-bag*, while in 2016 this rose to 23%. The difference between the two years is likely caused by a number of small bug-fixes applied to the *Book-bag*, which in 2016 made it more useful for the *goal-oriented* task.

The lack of any significant differences in the amount of time spent in the two main stages (*Explore* and *Search*) indicates that overall participants use of the system is consistent across the years. It is thus possible to use the two years to compare in detail how the two cohorts interacted with the same interface and task.

3 Cluster Analysis

The interface provided a lot of functionality within each of the three stages. In the initial analysis the focus will be on how frequently participants interacted with the available functions. The following actions were taken extracted from the log data:

- *browse* – the participant clicked on one of the topics in the tree in the *Explore* stage;

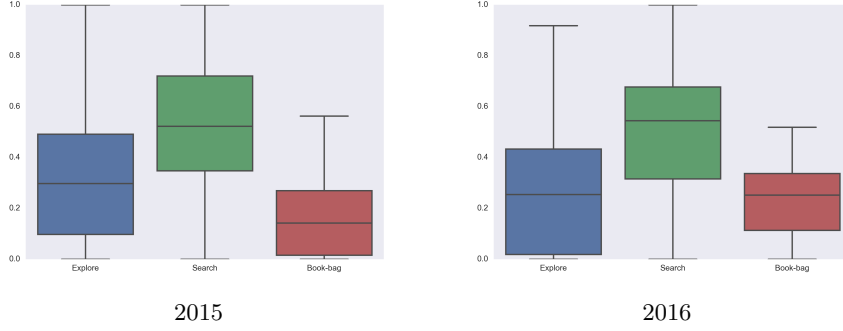


Fig. 3. Normalised time spent in each of the three stages of the multi-stage interface in the *goal-oriented* task.

- *query* – the participant issued a query, either by typing a query into the search box or by clicking on an item meta-data to search for that piece of meta-data;
- *facet* – the participant added or removed a facet in the *Search* stage;
- *paginate* – the participant paginated through the result list either in the *Explore* or *Search* stages;
- *item* – the participant viewed an item;
- *bookbag* – the participant interacted with the *Book-bag*, adding or removing a book, or adding notes to a book;
- *similar* – the participant used the similar items functionality in the *Book-bag*.

For each user a count of how often they used each of the actions was determined. The count vectors were then clustered using hierarchical, average-linkage clustering [5]. Clusters were determined using a distance threshold of 0.2. The resulting clusters were manually analysed and classified based on which actions participants in a cluster used. The cluster membership counts were normalised by the total number of participants in each year (2015: 95 both tasks; 2016: 52 *non-goal*, 53 *goal-oriented* task) to enable comparisons.

In neither year nor task did the *bookbag* and *similar* actions distinguish any of the clusters, as they are used relatively consistently across all clusters. The primary distinction between participants in both years was whether they primarily used the *browse* functionality in the *Explore* stage, or the *query* functionality in the *Search* stage, or whether they used a *mixed* approach (tab. 1 & 2). Those main distinctions were then augmented by distinctions based on what other actions they used. If an action is not included in the cluster name, then it was not distinctively used by participants in that cluster.

3.1 Non-goal Task

The cluster data in Table 1 mirrors the general time data from Figure 2, indicating that the majority of time is spent in the *Explore* stage (clusters# 1-4), with

varying levels of activity in the *Search* stage (cluster #5 and #6). The results also show that for this type of unstructured task the use of just the *query* action is not a common strategy (cluster #7: two users in 2015, none in 2016).

Table 1. Clusters extracted for the *non-goal* task. Cluster frequency normalised by total participant count (2015: 95 participants, 2016: 52 participants).

# Cluster	2015	2016
1 browse	.11 (10)	.00 (0)
2 browse + item	.68 (65)	.38 (20)
3 browse + paginate	.00 (0)	.04 (2)
4 browse + item + paginate	.00 (0)	.08 (4)
5 mixed	.06 (6)	.13 (7)
6 mixed + item	.07 (7)	.31 (16)
7 query	.02 (2)	.00 (0)
8 item	.02 (2)	.04 (2)
unclustered	.05 (2)	.02 (1)

Within those broad strokes, there are however quite significant differences between the two years. First, while in 2015 11% of participants only browsed and selected their books without looking at any item details (cluster #1), in 2016 no participant used this strategy. This fits into the larger picture, where from 2015 to 2016 there is a clear move from just browsing-based strategies (clusters #1-4) to strategies that incorporate a significant search element in the *mixed* strategies (clusters #5 and #6).

In both years there are two participants who basically did not interact with the system in any detail (cluster #8), but simply selected one browse topic, then scrolled through the results, and selected one or more books from that list into their book-bag. Further work is needed to investigate whether these users are generally disengaged from the task or whether they are just struggling with the open-ended nature of the *non-goal* task.

3.2 Goal-oriented Task

As with the *non-goal* task, the system use in the *goal-oriented* task (tab. 2) mirrors the time spent (fig. 3). Interestingly, the *goal-oriented* task shows larger clusters and more overlap between the years than the *non-goal* task.

The most striking difference between the two years is that in 2016 13% of participants used a pure browsing strategy to complete the *goal-oriented* task (cluster #1). This change comes at the expense of using a search-only strategy (cluster #5). There are also more participants who used a mixed strategy, but did not look at many items in detail (cluster #2). Nevertheless the differences between the two years are less than in the *non-goal* task, most likely as the focused nature of the task provides a guiding structure that reduces individual differences.

Table 2. Clusters extracted for the *goal-oriented* task. Cluster frequency normalised by total participant count (2015: 95 participants, 2016: 53 participants).

#	Cluster	2015	2016
1	Browse	.00 (0)	.13 (7)
2	Mixed	.05 (5)	.13 (7)
3	Mixed + Items	.61 (58)	.47 (25)
4	Mixed + Items + Facet	.02 (2)	.00 (0)
5	Query	.30 (28)	.21 (11)
	Unclassified	.02 (2)	.06 (3)

4 Browsing Patterns

Focusing on the use of browsing in the experiment, the second analysis focused on how participants interacted with the tree structure in the *Explore* stage. To facilitate the comparison of browsing patterns, the following browsing actions were extracted from the log data:

- *start* – the participant has not previously selected a topic and selects a top-level topic. This includes the scenario where the participant switches to the *Search* stage and then back;
- *depth* – the participant selects a child topic of the currently selected topic;
- *breadth* – the participant selects a sibling of the currently selected topic or a sibling of one of the current topic’s ancestors;
- *backtrack* – the participant selects one of the ancestor topics of the current topic;
- *restart* – the participant selects a top-level topic that is not related to the current topic.

In the analysis consecutive uses of the *depth*, *breadth*, and *backtrack* actions are merged into a single action before the analysis. This ensures that differences in the heights and breadths of individual sub-trees of the hierarchical topic structure do not influence the analysis. The resulting browse patterns were counted to identify all patterns that make up more than 5% of the total number of patterns (tab. 3 & 4).

A central result for both years and tasks is that the use of *backtrack* is limited to complex browse patterns that do not occur frequently. The maximum frequency for any browse pattern involving *backtrack* is 2%. However, overall *backtrack* occurs in approximately 7% - 12% of patterns.

4.1 Non-goal Task

In the *non-goal* task, as with the action use in section 3, there are significant differences between the two years. The main difference is that while in 2015 the most frequently used patterns are drilling down from the top with an optional

breadth-search at the bottom of the tree (pattern #3, #5, #7), in 2016 the use of looking at siblings at the bottom has reduced significantly (pattern #5), with a matching increase in straight drill-down behaviour (patterns #3 and #7). It seems that the 2016 participants saw significantly less value in looking around the leaves of a branch, preferring instead to go back to the top and delve into a different branch.

Table 3. Browse patterns for the *non-goal* task. Pattern frequency is normalised by the total number of patterns extracted from the log-data (2015: 173 patterns, 2016: 109 patterns). Patterns with less than 5% frequency have been filtered.

#	Pattern	2015	2016
1	start	.08	.06
2	start → breadth	.05	.05
3	start → depth	.14	.22
4	start → breadth → depth	.10	.08
5	start → depth → breadth	.15	.07
6	restart	.08	.05
7	restart → depth	.14	.24
8	restart → depth → breadth	.05	.06
9	restart → breadth → depth	.05	.00

4.2 Goal-oriented Task

The main difference is the big decrease in the frequency of just the *start* action (pattern #1). This mirrors the general increase in use of the browsing functionality (see fig. 2, cluster 1) in the *goal-oriented* task. This indicates that the 2016 participants saw significantly more benefit in exploring the collection using the *Explore* stage than in 2015. However, these participants only explored one branch in the tree, before then moving on to the *Search* stage, as evidenced by the lack of change in the *restart* patterns (#6-9). Further analysis is needed to investigate what the participants are learning from this single exploration.

The second interesting difference is that as in the *non-goal* task, there is a marked reduction in the number of uses of the depth-then-breadth patterns (#5 and #9). Why participants did not browse around the leaves of the branches requires further study.

5 Conclusion

Two years of iSBS data generated from the same data-set, tasks, and interface provide an interesting window into the potential variation between experiment participant cohorts. The overall comparison of the two years' of data indicates that the relative uses of the three main stages of the multi-stage interface are

Table 4. Browse patterns for the *goal-oriented* task. Pattern frequency is normalised by the total number of patterns extracted from the log-data (2015: 184 patterns; 2016: 99 patterns). Patterns with less than 5% frequency have been filtered.

#	Pattern	2015	2016
1	start	.27	.11
2	start → breadth	.02	.05
3	start → depth	.23	.27
4	start → breadth → depth	.09	.10
5	start → depth → breadth	.10	.03
6	restart → depth	.11	.11
7	restart → depth → breadth	.02	.05
8	restart → breadth	.03	.05
9	restart → breadth → depth	.06	.10

stable across the participant cohorts. Thus changes in detailed use patterns are likely due to individual differences, rather than due to external factors.

Both the analysis of the action use clusters in the system and the browse patterns reveal some quite large changes in detail behaviour between the two years. In particular there is a general trend towards using both the *Explore* and *Search* stages for both tasks. The 2016 data shows a decrease in the amount of pure-browsing approaches to the *non-goal* task together with an increase in the amount of browsing in the *goal-oriented* task.

This indicates that while for the general trends such as the overall use of the different stages, the number of participants in the two experiments (2015: 95, 2016: 105) is sufficient to produce stable and reliable results. However, for more detailed analysis the variation suggests that the results have to be read with a certain amount of caution, and potentially larger participant numbers are required to produce stable results.

Future work will need to investigate the data in more detail to determine whether it is possible to find explanations for the observed differences in the behaviour of the two cohorts. In particular the potential impact of cultural / language / background issues will need to be investigated.

References

1. Campbell, D., Hall, M.M., and Walsh, D.: Edge Hill Computing @ Interactive Social Book Search 2015. In: Cappellato, L., Ferro, N., Jones, G.J.F., and San Juan, E. (eds.) CLEF2015 Working Notes. CEUR Workshop Proceedings (2015)
2. Gäde, M., Hall, M.M., Huurdeman, H., Kamps, J., Koolen, M., Skov, M., Toms, E., and Walsh, D.: Overview of the SBS 2015 Interactive Track. In: Cappellato, L., Ferro, N., Jones, G.J.F., and San Juan, E. (eds.) CLEF2015 Working Notes. CEUR Workshop Proceedings (2015)
3. Hall, M.M., Fernando, S., Clough, P., Soroa, A., Agirre, E., and Stevenson, M.: Evaluating hierarchical organisation structures for exploring digital libraries. *Information Retrieval* 17(4), 351–379 (2014)

4. Hearst, M.A.: Search User Interfaces. Cambridge University Press (2009)
5. Jones, E., Oliphant, T., Peterson, P.: SciPy: Open source scientific tools for Python (2001–)
6. Kuhlthau, C.C.: Inside the search process: Information seeking from the user's perspective. *JASIS* 42(5), 361–371 (1991)
7. T.b.a: Overview of the SBS 2016 Interactive Track. In: CLEF2016 Working Notes. CEUR Workshop Proceedings (2016)
8. Vakkari, P.: A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of documentation* 57(1), 44–60 (2001)